

African Strains of Hepatitis E Virus That Are Distinct From Asian Strains

Ranee Chatterjee,¹ Sergei Tsarev,¹ Jacques Pillot,² Pierre Coursaget,³ Suzanne U. Emerson,¹ and Robert H. Purcell^{1*}

¹Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland

²Unité d'Immunologie Microbienne, Institut Pasteur, Paris, France

³Laboratoire d'Immunologie des Maladies Infectieuses, Faculté de Pharmacie Philippe Maupas, Tours, France

Partial genomic sequences of four hepatitis E virus (HEV) strains from Africa (Morocco and Tunisia) and one from Central Asia (Tashkent, Uzbekistan) were obtained. The reverse transcriptase-polymerase chain reaction was used to amplify 5' and hypervariable regions of open reading frame 1 (ORF1) and a region overlapping all 3 ORFs. Sequence analysis of these regions revealed the African strains to be quite distinct from all known Asian strains but more similar to them than to the Mexican strain. Sequence analysis of the Tashkent strain revealed almost complete identity with another central Asian strain from Osh, Kirgizia. These results thus further confirm the geographical origin of HEV strain divergence. *J. Med. Virol.* 53:139–144, 1997. © 1997 Wiley-Liss, Inc.[†]

KEY WORDS: enteric viral hepatitis; HEV sequence analysis

INTRODUCTION

Hepatitis E virus (HEV) is an enterically transmitted agent that causes epidemic and sporadic cases of hepatitis E, primarily in countries of Asia and Africa and in Mexico (Purcell and Ticehurst, 1988). It is a nonenveloped virus, 27–34 nm in diameter with a positive-strand RNA genome of 7.6 kb [Reyes et al., 1990; Bradley et al., 1987]. The genome is composed of 5' and 3' noncoding regions and 3 open reading frames (ORFs). The largest ORF, ORF1, located in the 5' region of the genome, encodes nonstructural proteins. ORF2, located in the 3' region, encodes the capsid protein, which elicits neutralizing antibodies. The very small ORF3 overlaps ORF1 and ORF2 and encodes an immunogenic protein of unknown function [Tam et al., 1991; Tsarev et al., 1994a]. A hypervariable region of approximately 300 nucleotides within ORF1 has been used to determine the genetic relatedness of strains [Tsarev et al., 1992; Huang et al., 1992].

The sequence of the full-length genome has been determined for HEV strains from Burma, Pakistan, Mexico and China [Tam et al., 1991; Aye et al., 1992, 1993; Tsarev et al., 1992; Huang et al., 1992; Yin et al., 1994]. In addition, the genomes of several Asian strains have been partially sequenced. Comparison of these sequences revealed a significant disparity between Asian strains and the Mexican strain. The Asian strains, while very similar to each other, can be divided into two subgroups, with the strains from Southeast Asia, including Burma and parts of India, in one subgroup and the strains from Northern and Central Asia, including Pakistan, China, Kirgizia, and India, in the other subgroup [Yin et al., 1994; Tsarev et al., 1992]. While HEV is prevalent in Africa, strains from this continent have not yet been characterized. In the present study, we partially sequenced the genomes of strains that were recovered from hepatitis E patients in Morocco and Tunisia and from one patient in Tashkent, Uzbekistan.

MATERIALS AND METHODS

Samples

Fecal (10% suspension in phosphate-buffered saline, pH 7.4) samples collected from patients with hepatitis during hepatitis E epidemics were inoculated intravenously into cynomolgus macaques. These samples included feces from a patient in Tashkent, Uzbekistan, three patients infected during a 1994 Moroccan epidemic and one patient from the Ivory Coast. In addition, bile from a cynomolgus monkey infected with HEV in feces obtained from a patient in Osh, Kirgizia during a 1988 epidemic was used to infect a rhesus monkey. An anti-HEV positive serum sample from a

Ranee Chatterjee's present address is Duke University School of Medicine, Durham, NC 27710.

Sergei Tsarev's present address is Walter Reed Army Institute of Research, Washington, DC 20307.

*Correspondence to: Robert H. Purcell, Bldg. 7, Rm. 202, NI-AID, 7 Center Drive MSC 0740, Bethesda, MD 20892-0740.

Accepted 30 May 1997

patient in Tunisia was available in very small quantity and, therefore, was not inoculated into a monkey.

Serum levels of liver enzymes [alanine aminotransferase (ALT) and isocitrate dehydrogenase (ICD)] and anti-HEV titers in the macaques were monitored weekly pre- and post-inoculation. Stool samples were collected thrice weekly and stored at -20°C . Serum samples were collected weekly and stored at -80°C . The facilities for the housing and care of the animals used in this experiment met or exceeded the requirements for primate husbandry.

Primers

Nested primers designed for amplification of the 5' and hypervariable regions of ORF1 and the ORFs1,2,3 overlap region were based on the published sequence of the SAR-55 Pakistani strain (GenBank accession number M80581). The following twelve primers corresponding to nucleotide positions of the M80581 sequence were used in this study. For the 5' region external sense primer (ESP) corresponds to nucleotide position 21-42, external antisense primer (EAP) corresponds to positions 443-421, internal sense primer (ISP) corresponds to positions 71-92 and internal antisense primer (IAP) corresponds to positions 353-334. For the hypervariable region the following primers were used: ESP (1830-1850), EAP (2424-2404), ISP (1957-1978), IAP (2364-2344). For the ORFs1,2,3 overlap region primers were: ESP (4959-4979), EAP (5297-5277), ISP (4966-4989), and IAP (5277-5254).

RNA Extractions

Thirty microliters of human serum or 100 μl of 10% stool suspensions from monkeys infected with the various HEV strains was extracted with (5:1) TRIzol reagent (Gibco BRL, Gaithersburg, MD) and chloroform. Isopropanol and glycogen were used for precipitation of RNA. The centrifuged pellet was then washed in 75% ethanol and dried.

cDNA Synthesis

Nucleic acid pellets were resuspended in 12 μl of water, and 1 μl of external antisense primer was added. The mixture was placed at 65°C for 5 minutes and then chilled on ice. Seven microliters of a master solution was added to yield final concentrations of 10 mM GeneAmp PCR buffer II (Promega, Madison, WI), 2.5 mM MgCl_2 (Promega), 1.25 mM dNTPs (Pharmacia Biotech, Gaithersburg, MD), 33U RNasin (Promega), and 10U AMV reverse transcriptase (Promega). The mixture was incubated in a 42°C water bath for 1 hour and then placed on ice.

PCR Reactions

First round amplification of the 5' region of ORF1 was performed by adding 20 μl of cDNA reaction mixture to 80 μl of a PCR reaction master solution containing 52.5 μl water, 8 μl each of GeneAmp 10 \times PCR buffer II and 25 mM MgCl_2 , 4 μl and 5 μl of antisense and sense primers respectively (7–8 μM), 0.5 mM dNTPs, and 2.5U AmpliTaq DNA polymerase (Perkin-Elmer, Oak Brook, IL).

First round amplification of the hypervariable region or overlap region was accomplished by adding the 20 μl of cDNA reaction mixture to an 80 μl solution containing 54 μl water, 10 μl 10 \times Pfu buffer (Stratagene, La Jolla, CA), 4% glycerol, 0.5 mM dNTPs, 4 and 5 μl respectively of antisense and sense primers, and 2.5U of Pfu cloned DNA polymerase (Stratagene).

Reaction mixtures were placed in thin-walled PCR tubes and overlaid with 100 μl of mineral oil. Amplification was carried out using a Perkin-Elmer Cetus DNA Thermal Cycler. The 5' region was amplified in 35 cycles consisting of 1 min at 94°C , 1 min at 45°C , and 1 min at 72°C . The hypervariable region and the overlap region were amplified for 35 cycles of 1 min at 94°C , 2 min at 42°C , and 2 min at 72°C .

Second round amplification was performed using 10 μl of first round PCR product in 90 μl of PCR mixture containing the internal antisense and sense primers. The cycling conditions were the same as for first round amplification. Products were visualized by UV light after electrophoresis in 1% agarose gel containing ethidium bromide.

DNA Purification and Cloning

Second round PCR-amplified DNA (30–35 μl reaction mixture) was electrophoresed on 1% agarose gel with ethidium bromide, and the appropriate bands were excised. The DNA was then purified with GeneClean or GeneClean Spin Filter kits (Bio 100).

Gel-purified PCR products of the 5' region and the overlap region were sequenced directly. The hypervariable region, however, was amplified much less efficiently and the amount of product was too low to provide the entire sequence from the gel-purified PCR product. This region was therefore cloned using the pCR-Script SK(+) cloning kit (Stratagene) according to the manufacturer's directions. The ligation reaction mixture was used to transform Epicurean Coli XL-1 Blue MRF'Kan supercompetent cells. Cells were plated on LB-ampicillin agar plates prepared with X-gal and IPTG and incubated overnight at 37°C . White colonies were chosen and grown in 5 ml of ampicillin-containing LB broth for 7–8 hours. Plasmids were purified with the QIAprep Spin Plasmid Kit (Qiagen, Chatsworth, CA).

Sequencing

Both strands of gel-purified PCR products and plasmid samples were sequenced, using the Applied Biosystems Inc. (ABI, Foster City, CA) Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase FS and ABI automated DNA sequencers, under a contract with the Frederick Cancer Research and Development Center, Frederick, MD. Sequences were deposited in GenBank under accession numbers: AF010417, AF010425 (Kirgizia); AF010418, AF010423, AF010429 (Mor/F12); AF010419 (Mor/F13); AF010420, AF010424, AF010428 (Mor/F23); AF010421 (Tunisia); AF010422, AF010426 (Uzbekistan); AF010427 (Tashkent). The consensus sequence was obtained for purified PCR products. The partial se-

TABLE I. Attempts to Transmit Hepatitis E Virus From Human Feces to Macaques

Inoculum	Origin	Hepatitis ^a (weeks)	Anti-HEV ^b (weeks)
1856	Uzbekistan (C. Asia)	Yes (8)	Yes (6)
F12	Morocco (Africa)	No	Yes (6)
F13	Morocco (Africa)	No	Yes (5)
F23	Morocco (Africa)	Yes (3)	Yes (3)
F163	Ivory Coast (Africa)	No	No

^aBiochemical evidence of hepatitis: ALT and/or ICD above base-line values.

^bSerological evidence of infection: seroconversion of IgG and IgM anti-HEV.

quences from uncloned PCR products as well as the entire sequence from three to four cloned products were compared to obtain the consensus sequence of the hypervariable region. Nucleotide and amino acid sequences of HEV strains were compared using the "GCG" software package, version 8.0 on a VAX 8650 computer (NIH, National Cancer Institute, Frederick, MD). Phylogenetic trees were constructed by the unweighted pair-group method with arithmetic mean that is based on the assumption of a constant rate of evolution.

RESULTS

The fecal samples were not sufficient to determine infectious viral titer and would be inadequate for extended studies. Therefore, transmission of HEV to macaque monkeys was performed in order to amplify the virus strains. Four of the five fecal samples (1856, F12, F13, F23) contained infectious virus and transmitted hepatitis E virus as measured by the appearance of IgG and IgM anti-HEV (Table I). The interval from inoculation to appearance of anti-HEV ranged from 3 to 6 weeks; IgG and IgM class antibodies were first detected during the same week in each case. Only two of the four monkeys (inoculated with 1856 and F23) infected with HEV from human feces developed biochemical evidence of hepatitis. The incubation period to onset of hepatitis was 8 weeks in the monkey inoculated with feces from Uzbekistan and 3 weeks in a monkey inoculated with one of the Moroccan fecal samples. The longer incubation period for the Uzbekistan compared to the Moroccan strain could reflect a lower titer of infectious virus since the length of time to seroconversion varied inversely with dose in experimentally infected cynomolgus macaques [Tsarev et al., 1994b]. The fecal sample from the Ivory Coast failed to transmit HEV or hepatitis to monkeys, despite two attempts. A rhesus monkey inoculated with bile from a cynomolgus monkey previously infected with HEV from Kirgizia developed hepatitis and IgG and IgM anti-HEV 3 weeks after inoculation (data not shown).

Although all three monkeys inoculated with Moroccan strains from the same epidemic were infected, ORF1 5' and hypervariable sequences could be amplified from the feces of only two of these (F12 and F23). However, the region overlapping all three ORFs was

amplified for all three strains. These three Moroccan strains had almost identical sequences (Tables II–IV). Comparison of the sequences from the 5' end of ORF1 revealed only 1 nucleotide difference, a C substituted for a T, resulting in a silent mutation. Comparison of the hypervariable region revealed a 99% identity with two C-for-T substitutions, both of which were also silent mutations. The F13 and F23 strains were identical in the overlap region and differed from strain F12 by a single nucleotide which resulted in an amino acid change.

Sequence from the Tunisian strain was obtained only from the overlap region, the most conserved of the three regions sequenced. The Tunisian strain was most homologous to the Moroccan strains (96%) but was only slightly less homologous (94–95%) to the Asian strains (Table IV). Its degree of homology with the Mexican strain (86%) was similar to that of all the other strains (84–87%).

Two different samples were used to obtain the sequence of the HEV strain from Tashkent, Uzbekistan (Tables II–IV). The first sample was a fecal sample obtained directly from the patient, from which only the hypervariable region could be amplified. The second sample consisted of fecal material collected from a rhesus monkey that had been inoculated with the patient's feces. The sequence of the hypervariable region from these two samples differed by two nucleotides, both of which were C-for-A missense mutations that resulted in a 98% amino acid identity.

Comparison of the strain from Tashkent, Uzbekistan with a second central Asian strain from Osh, Kirgizia revealed 99%, 98%, and 100% nucleotide identities in the 5', the hypervariable and the overlap regions, respectively. The amino acid sequences were 100%, 97%, and 100% identical in the three regions.

These sequences were then compared to HEV sequences listed in Genbank. Comparison of the sequence of the Moroccan strains with other HEV sequences revealed differences in all three regions, with the difference in the hypervariable region being most pronounced (Tables II–IV). Comparison of the Moroccan sequences with those of the two most distant Asian strains, the Pakistani and Burmese strains, revealed a nucleotide identity of 92% and 93% respectively in the 5' region, 79% and 80–81%, respectively in the hypervariable region and 95% and 95–96%, respectively in the overlap region. These identities are generally less than those between any two Asian strains, which share 93–99% identity in the 5' region, and 89–98% identity in the hypervariable region and 96–100% identity in the overlap region. Comparison of the Moroccan strains with the Mexican strain revealed nucleotide identities of 81, 52, and 86%, respectively, for the 5', hypervariable and overlap region similar to those between Asian and Mexican strains. Phylogenetic trees of the 5' and hypervariable regions placed the Moroccan strains on a branch located off of the main branch that contains the various Asian strains, while the Mexican branch remained very distinct (data not shown). The same gen-

TABLE II. Interstrain Comparisons of the 5' End of ORF1

HEV strain	Osh	Tash	KS2 ^a	SAR ^b	B2 ^c	F23	F12	Mex ^d
Nucleotides (243 total)								
Osh		3 ^e	14	17	12	16	17	50
%		99 ^f	94	93	95	93	93	79
Tash	0 ^g		11	14	11	17	18	51
%	100 ^h		95	94	95	93	93	79
KS2	1	1		3	12	16	17	47
%	99	99		99	95	93	93	81
SAR	1	1	0		15	19	20	50
%	99	99	100		94	92	92	79
B2	2	2	2	2		16	17	49
%	98	98	98	98		93	93	80
F23	1	1	1	1	1		1	47
%	99	99	99	99	99		100	81
F12	1	1	1	1	1	0		47
%	99	99	99	99	99	100		81
Mex	5	5	5	5	4	4	4	
%	94	94	94	94	95	95	95	
Amino Acids (81 total)								

^aKS2 = Kashi, China, 1987.

^bSAR = Sargodha, Pakistan, 1987.

^cB2 = Rangoon, Burma, 1986.

^dMex = Telixtac, Mexico, 1986.

^eNucleotide differences.

^f% nucleotide identity.

^gAmino acid differences.

^h% amino acid identity.

eral pattern was observed for the overlap region and additionally, the sample from Tunisia grouped with the Moroccan strains (Fig. 1).

Comparison of the two central Asian strains with the two prototype Asian strains showed that although they were equally related to the Burmese and Pakistani strains in the overlap region (97%; Table IV), they were more closely related to the Burmese strain than to the Pakistani strain in the 5' region of ORF1 (range of 93–95%; Table II) but the opposite was true in the hypervariable region (range of 89–93%; Table III). Both central Asian strains showed a 79% identity with the Mexican strain in the 5' region, a 52–53% identity in the hypervariable region, and 86% identity in the overlap region (Tables III–V).

DISCUSSION

This study examined the sequences of two conserved regions and the hypervariable region of four African isolates and two central Asian strains of HEV. The three samples obtained from patients in the same Moroccan epidemic displayed 99–100% amino acid and nucleotide identity in all three of the genomic regions examined. These data support the theory that HEV has a relatively stable genome.

Analysis of each region of the Moroccan strains revealed sequence divergence from other known strains and suggested a genetic relationship closer to the Asian strains than to the Mexican strain (Fig. 1). Unfortunately, only the sequence of the overlap region could be obtained for the Tunisian strain. However, based on sequences from this limited region, the Tunisian strain

clearly grouped with the Moroccan strains. Comparison of the strains from Osh, Kirgizia, and Tashkent, Uzbekistan indicated a very close relationship between the two, with nearly identical nucleotide and amino acid sequences. Analysis of the three genomic regions examined in this study, as well as of a region of the RNA polymerase, which had been determined previously for the Osh strain [Tsarev et al., 1992], indicated that these Central Asian strains were located genetically between the prototype Pakistani and Burmese strains.

The sequences of the Moroccan and central Asian strains of HEV clearly reflect an evolution of HEV strains based on geographical origin. The extreme divergence of the Mexican strain, coupled with its circulation in the hemisphere opposite that of the Asian and African strains, also supports this pattern. It will be interesting to determine if new strains can be isolated in other parts of the world and whether analyses of them will clarify the relationships among the presently known strains.

ACKNOWLEDGMENTS

We thank Marianne Lewis and Tatiana Tsareva for technical assistance, T. Heller, Y. Huang, G. Raychaudhuri, and X-J Meng for their input and guidance, and the Duke University School of Medicine for allowing this study project to be accomplished.

We acknowledge the National Cancer Institute for allocation of computing time and staff support at the Advanced Scientific Computing Laboratory of the Frederick Cancer Research Development Center. This

TABLE III. Interstrain Comparisons of the Hypervariable Region of ORF1

HEV strain	Osh	Tash	KS2 ^a	SAR ^b	B2 ^c	F23	F12	Mex ^d
Nucleotides (321 total)								
Osh		6 ^e	26	25	36	65	65	153
%		98 ^f	92	92	89	80	80	52
Tash	3 ^g		22	21	32	66	66	151
%	97 ^h		93	93	90	79	79	53
KS2	12	12		9	31	67	67	153
%	89	89		97	90	79	79	52
SAR	14	14	4		32	66	66	151
%	87	87	96		90	79	79	53
B2	16	16	16	18		62	63	148
%	85	85	85	83		81	80	54
F23	22	23	23	25	23		2	155
%	79	78	78	77	78		99	52
F12	22	23	23	25	23	0		154
%	79	78	78	77	78	100		52
Mex	65	65	67	65	64	65	65	
%	39	39	37	39	40	39	39	
Amino Acids (107 total)								

^aKS2 = Kashi, China, 1987.^bSAR = Sargodha, Pakistan, 1987.^cB2 = Rangoon, Burma, 1986.^dMex = Telixtac, Mexico, 1986.^eNucleotide differences.^f% nucleotide identity.^gAmino acid differences.^h% amino acid identity.

TABLE IV. Interstrain Comparisons of the ORFs1,2,3 Overlap Region

HEV strain	Osh	Tash	KS2 ^a	SAR ^b	B2 ^c	F13	F23	F12	Tg60 ^d	Mex ^e
Nucleotides (232 total)										
Osh		0 ^f	6	6	7	12	12	11	12	33
%		100 ^g	97	97	97	95	95	95	95	86
Tash	0 ^h		6	6	7	12	12	11	12	33
%	100 ⁱ		97	97	97	95	95	95	95	86
KS2	2	2		4	10	14	14	13	14	32
%	98	98		98	96	94	94	94	94	86
SAR	0	0	2		10	12	12	11	12	29
%	100	100	98		96	95	95	95	95	87
B2	0	0	2	0		11	11	10	11	36
%	100	100	98	100		95	95	96	95	84
F13	2	2	4	2	2		0	1	9	32
%	98	98	96	98	98		100	100	96	86
F23	2	2	4	2	2	0		1	9	32
%	98	98	96	98	98	100		100	96	86
F12	1	1	3	1	1	1	1		10	33
%	99	99	97	99	99	99	99		96	86
Tg60	4	4	6	4	4	4	4	4		32
%	96	96	96	96	96	96	96	96		86
Mex	12	12	13	12	12	12	11	12	14	
%	90	90	89	90	90	90	90	90	88	
Amino Acids (116 total)										

^aKS2 = Kashi, China, 1987.^bSAR = Sargodha, Pakistan, 1987.^cB2 = Rangoon, Burma, 1986.^dTg60 = Tunis, Tunisia, 1986.^eMex = Telixtac, Mexico.^fNucleotide differences.^g% nucleotide identity.^hAmino acid differences.ⁱ% amino acid identity.

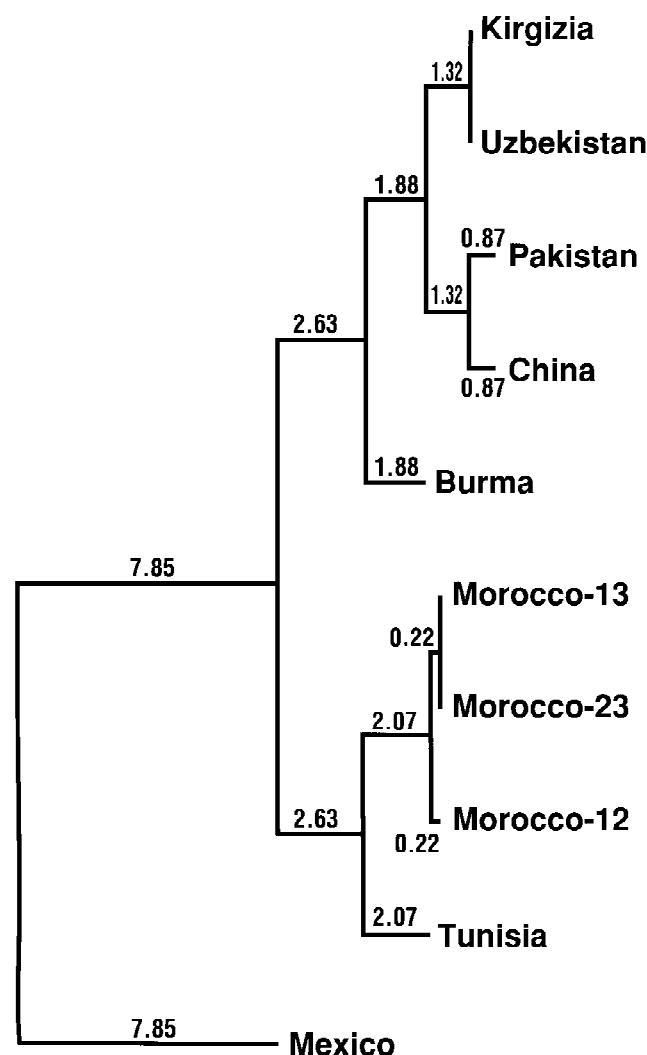


Fig. 1. Phylogenetic tree of the genetic relationships of HEV isolates based on nucleotide sequences from a region overlapping ORFs 1, 2, and 3. Numbers on the branches represent percent differences.

study was supported in part by contracts N01-AI-45180 and N01-AI-52705.

REFERENCES

- Aye TT, Uchida T, Ma X-Z, Iida F, Shikata T, Zhuang H, Win KM (1992): Complete nucleotide sequence of a hepatitis E virus isolated from the Xinjiang epidemic (1986-1988) of China. *Nucleic Acids Research* 20:3512.
- Aye TT, Uchida T, Ma X-Z, Iida F, Shikata T, Ichikawa M, Rikihisa T, Win KW (1993): Sequence and gene structure of the hepatitis E virus isolated from Myanmar. *Virus Genes* 7:95-110.
- Bradley DW, Kravczynski K, Cook EH, McCaustland KA, Humphrey CD, Spelbring JE, Myint H, Maynard JE (1987): Enterically-transmitted non-A, non-B hepatitis: Serial passage of disease in cynomolgus macaques and tamarins and recovery of disease-associated 27 to 34 nm virus-like particles. *Proceedings of the National Academy of Sciences USA* 84:6277-6281.
- Huang C-c, Nguyen D, Fernandez J, Yun KY, Fry KE, Bradley DW, Tam AW, Reyes GR (1992): Molecular cloning of the Mexico isolate of hepatitis E virus (HEV). *Virology* 191:550-558.
- Purcell RH, Ticehurst JR (1988): Enterically transmitted non-A, non-B hepatitis: Epidemiology and clinical characteristics. In Zuckerman AJ (ed): "Viral Hepatitis and Liver Disease." New York: Alan R. Liss, pp 131-137.
- Reyes GR, Purdy MA, Kim JP, Luk K-C, Young LM, Fry KE, Bradley DW (1990): Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* 247:1335-1339.
- Tam AW, Smith MM, Guerra ME, Huang C-C, Bradley DW, Fry KE, Reyes GR (1991): Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome. *Virology* 185:120-131.
- Tsarev SA, Emerson SU, Reyes GR, Tsareva TS, Legters LJ, Malik IA, Iqbal M, Purcell RH (1992): Characterization of a prototype strain of hepatitis E virus. *Proceedings of the National Academy of Sciences USA* 89:559-563.
- Tsarev SA, Tsareva TS, Emerson SU, Govindarajan S, Shapiro M, Gerin JL, Purcell RH (1994a): Successful passive and active immunization of cynomolgus monkeys against hepatitis E. *Proceedings of the National Academy of Sciences USA* 91:10198-10202.
- Tsarev SA, Tsareva TS, Emerson SU, Yarbough PO, Legters LJ, Moskal T, Purcell RH (1994b): Infectivity titration of a prototype strain of hepatitis E virus in cynomolgus monkeys. *Journal Medical Virology* 43:135-142.
- Yin S, Purcell RH, Emerson SU (1994): A new Chinese isolate of hepatitis E virus: Comparison with strains recovered from different geographical regions. *Virus Genes* 9:23-32.